

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:09:31 ; Search time 299.73 Seconds  
(without alignments)  
18.379 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_33\_79

Perfect score: 275  
Sequence: 1 GDDPPKRYEDCRRRCCEWDT.....QCESCKSYGKEDQDQQRH 47

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL\_15:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	124	45.1	525	10 Q43358	Q43358 theobroma c
2	105.5	38.4	593	10 Q9SEW4	Q9SEW4 juglans reg
3	94	34.2	666	10 Q9SPL4	Q9SPL4 macadamia i
4	92	33.5	666	10 Q9SPL5	Q9SPL5 macadamia i
5	90	32.7	810	10 Q9ZWT3	Q9ZWT3 cucurbita m
6	88	32.0	625	10 Q9SPL3	Q9SPL3 macadamia i
7	82.5	30.0	393	10 Q9ZTP0	Q9ZTP0 oryza sativ
8	67.5	24.5	411	5 Q9V7P3	Q9V7P3 drosophila
9	65.5	23.8	554	5 P91419	P91419 caenorhabdi
10	65	23.6	554	5 Q9VPS3	Q9VPS3 drosophila
11	64	23.3	1298	4 Q9P2D9	Q9P2D9 homo sapien
12	63.5	23.1	242	10 Q81260	Q81260 tripsacum d
13	63.5	23.1	246	10 Q81261	Q81261 tripsacum d
14	62.5	22.7	103	11 Q9JIK0	Q9JIK0 rattus norv
15	62	22.5	525	5 Q9VSC2	Q9VSC2 drosophila
16	61.5	22.4	238	10 Q81255	Q81255 zea mays su
17	61	22.2	204	5 O16405	O16405 caenorhabdi
18	60.5	22.0	160	10 Q9S0H1	Q9S0H1 atarchis hyp
19	60.5	22.0	238	10 Q81257	Q81257 zea luxuria

20	60.5	22.0	396	4 Q9NSJ1	Q9NSJ1 homo sapien
21	60.5	22.0	897	5 Q17336	Q17336 caenorhabdi
22	60.5	22.0	1327	5 Q9NFB4	Q9NFB4 plasmodium
23	60.5	22.0	1819	2 Q9ZLIV0	Q9ZLIV0 helicobacte
24	60.5	22.0	3469	5 Q9U4I2	Q9U4I2 drosophila
25	60	21.8	241	2 Q9K1Y2	Q9K1Y2 chlamydia p
26	60	21.8	456	2 Q9Z9G3	Q9Z9G3 chlamydia p
27	60	21.8	774	2 Q9K1Z4	Q9K1Z4 chlamydia p
28	60	21.8	774	2 Q9JSK8	Q9JSK8 chlamydia p
29	60	21.8	811	2 Q9Z6K5	Q9Z6K5 chlamydia p
30	60	21.8	835	5 Q9V744	Q9V744 drosophila
31	60	21.8	1927	2 Q25262	Q25262 helicobacte
32	59.5	21.6	122	10 Q03863	Q03863 zea mays (m
33	59.5	21.6	163	12 Q9Q098	Q9Q098 chimpanzee
34	59.5	21.6	191	6 Q9N0L8	Q9N0L8 macropus eu
35	59.5	21.6	230	2 Q9X1D9	Q9X1D9 thermotoga
36	59.5	21.6	236	10 Q81250	Q81250 zea mays su
37	59.5	21.6	236	10 Q9SBE2	Q9SBE2 zea mays su
38	59.5	21.6	238	10 Q81249	Q81249 zea mays su
39	59.5	21.6	238	10 Q81258	Q81258 zea luxuria
40	59.5	21.6	238	10 Q9SBE8	Q9SBE8 zea mays su
41	59.5	21.6	238	10 Q9SBE5	Q9SBE5 zea luxuria
42	59.5	21.6	238	10 Q9S6E6	Q9S6E6 zea luxuria
43	59.5	21.6	239	10 Q9SBE1	Q9SBE1 zea mays su
44	59.5	21.6	240	10 Q81252	Q81252 zea mays su
45	59.5	21.6	240	10 Q81256	Q81256 zea mays su

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	525 AA.
ID	Q43358			
AC	Q43358:			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	VICILIN PRECURSOR.			
GN	CSV.			
OS	Theobroma cacao (Cacao).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Malvales; Malvaceae; Theobroma.			
OX	NCBI_TaxID=3641;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAVES;			
RX	MEDLINE=92288309; PubMed=1600151;			
RA	McHenry L., Filtz P.J.;			
RT	*Comparison of the structure and nucleotide sequences of vicilin genes			
RT	of cocoa and cotton raise questions about vicilin evolution.*;			
RL	Plant Mol. Biol. 18:1173-1176(1992).			
DR	EMBL: X62625; CAA44493.1; -;			
DR	EMBL: X62626; CAA44494.1; -;			
DR	HSSP: P02853; 2PHL.			
DR	MENDEL: 30919; Thecc:1188;30919.			
DR	INTERPRO: IPR01113; -;			
DR	PFAM: PF00546; Seedstore_7s; 1.			
KW	PRODOM: PD081059; -; 1.			
DR	Signal.			
FT	SIGNAL.			
FT	CHAIN			
FT	CHAIN			
FT	SEQUENCE			

Query Match 45.1%; Score 124; DB 10; Length 525;  
Best Local Similarity 46.3%; Pred. No. 5.2e+08;  
Matches 19; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 4 DPKRYEDCRRRCCEWDTROKQEDQCESCKSYGKEDQDQ 44  
DB 37 DPROGYECORRCSEATEEREDEOCRCREREYKEDQDQ 77

RESULT 2  
Q9SEW4 PRELIMINARY: PRT: 593 AA.  
AC Q9SEW4: 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE VICILIN-LIKE PROTEIN PRECURSOR (FRAGMENT).  
OS Juglans regia (English walnut).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid I;  
OC Fagales; Juglandaceae; Juglans.  
OX NCBI\_TaxID=51240;  
GN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. SUNLAD; TISSUE=SONATIC EMBRYO LINE;  
RA Teuber S.S., Jarvis K.C., Peterson W.R., Dandekar A.M., Ansari A.A.;  
RT "Identification and cloning of a cDNA encoding a vicilin-like protein,  
RT Jug r 2, from English walnut kernel (Juglans regia): a major food  
RT allergen.";  
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF066055; AAF18269.1; -  
DR HSSP: P02853; 2PHL.  
DR INTERPRO: IPR001113; -  
DR PFAW: PF00546; Seedstore\_7s: 1.  
FT NON\_TER 1  
SQ SEQUENCE 593 AA: 69990 MW; 9BA127E19B18C0D8 CRC64;

Query Match 38.4%; Score 105.5; DB 10; Length 593;  
Best Local Similarity 29.0%; Pred. No. 1.3e-05;  
Matches 18; Conservative 13; Mismatches 12; Indels 19; Gaps 1;  
QY 4 DPPPKRYEDCRRRCCEMDTRGOKEQOOCCEESCKSQY-----GEMDQOQ 44  
DB 118 DDPQTHRCQRCQIOEOSPERRQCQRCERQYKQGRERPEASPRRESRGREEO 177  
OY 45 RH 46  
DB 178 RH 179

RESULT 3  
Q9SPL4 PRELIMINARY: PRT: 666 AA.  
AC Q9SPL4: 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE VICILIN PRECURSOR.  
GN AMP2.  
OS Macadamia integrifolia (Macadamia nut).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.  
OX NCBI\_TaxID=60698;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=NUT KERNEL;  
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;  
RT "A family of antimicrobial peptides is produced by processing of a 7S  
RT globulin protein in Macadamia integrifolia.";  
RL Plant J. 0:0-0(1999).  
DR EMBL: AF161884; AAD54245.1; -  
DR HSSP: P02853; 2PHL.  
DR INTERPRO: IPR001113; -  
DR PFAW: PF00546; Seedstore\_7s: 1.  
SQ SEQUENCE 666 AA: 78243 MW; 0ECA22F8710F8A7B CRC64;

Query Match 34.2%; Score 94; DB 10; Length 666;  
Best Local Similarity 35.6%; Pred. No. 0.00042;

Matches 16; Conservative 8; Mismatches 21; Indels 0; Gaps 0;  
QY 3 DPPPKRYEDCRRRCCEMDTRGOKEQOOCCEESCKSQYGEKDQOORH 47  
DB 76 DDPQTHRCQRCQIOEOSPERRQCQRCERQYKQGRERPEASPRRESRGREEO 120

RESULT 4  
Q9SPL5 PRELIMINARY: PRT: 666 AA.  
AC Q9SPL5: 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE VICILIN PRECURSOR.  
GN AMP2.  
OS Macadamia integrifolia (Macadamia nut).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.  
OX NCBI\_TaxID=60698;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=NUT KERNEL;  
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;  
RT "A family of antimicrobial peptides is produced by processing of a 7S  
RT globulin protein in Macadamia integrifolia kernels.";  
RL Plant J. 0:0-0(1999).  
DR EMBL: AF161883; AAD54244.1; -  
DR HSSP: P02853; 2PHL.  
DR INTERPRO: IPR001113; -  
DR PFAW: PF00546; Seedstore\_7s: 1.  
SQ SEQUENCE 666 AA: 78217 MW; C752B884B2DF0224 CRC64;

Query Match 33.5%; Score 92; DB 10; Length 666;  
Best Local Similarity 33.3%; Pred. No. 0.00075;  
Matches 15; Conservative 9; Mismatches 21; Indels 0; Gaps 0;  
QY 3 DPPPKRYEDCRRRCCEMDTRGOKEQOOCCEESCKSQYGEKDQOORH 47  
DB 76 DDPQTHRCQRCQIOEOSPERRQCQRCERQYKQGRERPEASPRRESRGREEO 120

RESULT 5  
Q9ZWI3 PRELIMINARY: PRT: 810 AA.  
AC Q9ZWI3: 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE PV100.  
OS Cucurbita maxima (Pumpkin) (Winter squash).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid I;  
OC Cucurbitales; Cucurbitaceae; Cucurbita.  
OX NCBI\_TaxID=3661;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KURUKAWA AMAKURI NANKIN; TISSUE=COTYLEDON;  
RX MEDLINE=99107919; PubMed=9891029;  
RA Yamada K., Shimada T., Kondo M., Nishimura M., Hara-Nishimura I.;  
RT "Multiple functional proteins are produced by cleaving Asn-Gln bonds  
RT of a single precursor by vacuolar processing enzyme.";  
RL J. Biol. Chem. 274:2563-2570(1999).  
DR EMBL: AB019195; BAA34056.1; -  
DR HSSP: P02853; 2PHL.  
DR INTERPRO: IPR001113; -  
DR PFAW: PF00546; Seedstore\_7s: 1.  
DR PRODOM: PD081059; -; 1.  
SQ SEQUENCE 810 AA: 97314 MW; A829A3F7542266AB CRC64;

Query Match 32.7%; Score 90; DB 10; Length 810;

Best Local Similarity 39.5%; Pred. No. 0.0016;  
Matches 15; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

```

QY      5 PPKRRDCCRRCCEMDTRGQKEDQCCCEFSKCSQYGEKDQ 42
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      74 PRAVEYVCRRLCQVAERGVGEQQKCEQYCEERLRERQ 111

```

RESULT	6
Q9SPL3	
ID	Q9SPL3
PRELIMINARY;	
PRT;	625 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, last annotation update),  
DE VICILIN PRECURSOR (FRAGMENT).

OS *Macadamia integrifolia* (Macadamia nut).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta  
CC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.  
OX NCBI\_TaxID=60698;

RP SEQUENCE FROM N.A.  
RC TISSUE=NTU KERNEL:  
RA Marcus J.P., Coulter K.C., Green J.L., Manners J.M.;  
RT "A family of antimicrobial peptides is produced by pro-  
RT globulin protein in *Macadamia integrifolia*.";  
RL Plant J. 0:0-0(1999).  
DR EMBL: AF161885; AAD54246.1; -.  
DR HSSP: P02853; 2PHL.  
DR INTERPRO: IPR001113. -.  
DR Pfam: PF00546; Seedstore\_7s; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE .625 AA; 73586 MW; 415808A89D370296 CRC64;

Query Match	32.08	Score 88	DB 10	Length 625
Best Local Similarity	29.58	Pred. No. 0.0023		
Matches 13	Conservative 15	Mismatches 16	Indels 0	Gaps 0

QY 4 DPKRYEDCRRRCCEWDTRGCKEQQOCEESCKSOYGEKKDQQQRHR 47  
 ||::|||::||: ||::||: ||::||: |  
 Db 80 DPGQGYEGCCQRCQRRRETEPRHMQICQRCERRYEKEKKRQQR 123

RESULT	7	
Q9ZTP0		
ID	Q9ZTP0	PRELIMINARY;
		PRT; 393 AA.

DT	01-MAY-1999	(TREMblrel, 10, Created)
DT	01-MAY-1999	(TREMblrel, 10, last sequence update)
DT	01-MAY-1999	(TREMblrel, 10, last annotation update)
DE	HYPOTHETICAL	45.3 KDA PROTEIN.

OS *Oryza sativa* (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
 OX NCBI\_TaxID=4530;

RP SEQUENCE FROM N.A.  
RC STRAIN=LOMELLO;  
RA Chen P.W., Chen L.J.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases

SEQ	SEQUENCE	393	AA;	45258	MM;	DBD01934BA2F9E95	CRC64;
-----	----------	-----	-----	-------	-----	------------------	--------

Matches 14; Conservative 13; Mismatches 19; Indels 2

1 GDDDPKRYEDCRRCEWDT-RGQKEQQQCEESC--KSQYGEKDDQQRH 46

```

Db 32 GRDPKEELRWCKKQCRWEAGDQRLRECEEQCLQROEDDDDDENTH 80

```

RESULT	8
Q9V7P3	
ID	Q9V7P3
PRELIMINARY;	
PRT;	541 AA

DT	01-MAY-2000	(TREMblrel. 13, Created)
DT	01-MAY-2000	(TREMblrel. 13, Last sequence update)
DT	01-MAY-2000	(TREMblrel. 13, Last annotation update)
DE	CG6262	PROTEIN.

05 *Drosophila melanogaster* (Fruit fly).  
0C Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;  
0C Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
0C Ephyraoidea; Drosophilidae; *Drosophila*.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;

RA Ananatlides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA April J., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
BA Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
BA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
BA Borison D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
BA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz-Costa S., Ferrelira S., Fleischmann W.,

RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
JA Jalali M., Kalish F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merklov G., Milshina N.V., Mobbary C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Rount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,  
RA Shue B.C., Siden-kiomos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA

RA Wang Z.-Y., Waststman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
ZHANG X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
ZHANG X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003807; AAF58004.1; -  
 FLYBASE: FBqnd0034121; CG6262.

Query Match	24.58;	Score 67.5;	DB 5;	Length 541;
Best Local Similarity	39.68;	Pred. No. 0.8;		

```

234 DTDP-----DC--EC$MDEKQKQMEQKQKQDCMRQ--Q000002EHQ 2722

```

```

RESULT 9
ID P91419 PRELIMINARY: PRT: 411 AA.
AC P91419:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE CODD FOR BY C. ELEGANS CDNA YK15A6.5.
GN T01D1.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Felodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2.
RA Bradshaw H., Wohlmann P.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2.
RA Waterston R.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U80455; AAB37887.1; -;
SQ SEQUENCE 411 AA; 44675 MW; 599DAC9DCFA1382 CRC64;

Query Match 23.8%; Score 65.5; DB 5; Length 411;
Best Local Similarity 35.9%; Pred. No. 1.1;
Matches 14; Conservative 11; Mismatches 11; Indels 3; Gaps 2;

OY 10 EDCRRRCWMDTRGQKE-QOOCESCKS--QYGEKDOQR 45
DB 147 OPCQACPPQOQPOOQOCCOTCQSDPDQYSQOLIQQ 185

RESULT 10
O9VPS3 PRELIMINARY: PRT: 554 AA.
AC O9VPS3:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE CG2839 PROTEIN (FRAGMENT).
GN CG2839.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC NCB1_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Meruliov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relneert K., Remington K., Saunders R.D.C., Schelder F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Wellstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003588; AAF51469.1; -;
DR FLYBASE: FBgn0031273; CG2839.
DR INTERPRO: IPR001304; -;
DR PFAM: PF00059; lectin_G_1.
DR PROSITE: PS50041; C_Type_Lectin_2; 1.
FT NON_TER 554 554
SQ SEQUENCE 554 AA; 69873 MW; 029B6D852609FD16 CRC64;

Query Match 23.6%; Score 65; DB 5; Length 554;
Best Local Similarity 35.9%; Pred. No. 1.7;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

OY 7 KRYEDCRRRCWMDTRGQKEQOOCESCKSQYGEKDOQR 45
DB 468 RRREKRKRREERREKRREERREERREKRREERKRKDERR 506

RESULT 11
O9P2D9 PRELIMINARY: PRT: 1298 AA.
AC O9P2D9:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE KIA1408 PROTEIN (FRAGMENT).
GN KIA1408.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Nagase T., Kikuno R., Ishikawa K., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.

```



